

1 Introduction

Vous êtes missionnés auprès de la banque *RASPETOUT* pour développer un modèle prédictif qui permettra d'établir des prévisions du nombre de clients se rendant au guichet de son agence située au 1 rue Bernard Lawrence Madoff à Tours. L'extrême variabilité de ce flux de clients se rendant physiquement à l'agence pose une double difficulté opérationnelle à la banque.

D'une part déterminer à l'avance le nombre d'employés qu'il faudrait affecter au guichet en fonction de l'affluence. Le risque est d'affecter trop de personnel au guichet au détriment des gestionnaires en back-office qui eux sont débordés de travail...

D'autre part s'assurer que le délai d'attente des clients, avant prise en charge par un conseiller de la banque, n'est pas trop long. Le risque est de susciter une très grande insatisfaction d'autant que la clientèle Tourangelle a de fortes exigences en terme de qualité de service.

En ce sens votre prévision, du nombre de visiteurs, va permettre à l'agence d'anticiper l'affluence, de s'organiser en terme de capacité et de planning des agents affectés au guichet et par ailleurs de mieux satisfaire ses clients.

2 Documents et données fournies

2.1 Base de données

Description

Données quotidiennes du 02/01/2010 au 31/05/2018

nombre d'observations : 2550

Nom du fichier `CrowdCounderBDD`

La base de données contient les variables suivantes:

`nb` nombre de clients présents à l'agence un jour donné

`t` Indice du mois

`date` la date

`jour` le rang du jour dans le mois

`mois` le rang du mois dans l'année

`an` l'année

`joursem` le jour de la semaine

`vacances` indicateur qui vaut 1 si le jour fait partie de la période des vacances scolaires

Exemple, extrait des 3 premières lignes du tableau de données

| nb | t | date | jour | mois | an | joursem | vacances |
|------|---|------------|------|------|------|---------|----------|
| 462 | 0 | 02/01/2010 | 2 | 1 | 2010 | Samedi | 0 |
| 1021 | 0 | 04/01/2010 | 4 | 1 | 2010 | Lundi | 0 |
| 735 | 0 | 05/01/2010 | 5 | 1 | 2010 | Mardi | 0 |

2.2 script R Markdown

script pour les estimations et l'output de l'analyse au forma html :
DevoirEconometrieNom1Nom2crowdcounter.Rmd

3 Question

Il s'agit de construire et d'estimer un modèle qui prédit au mieux le flux des clients quotidiens à l'agence.

Il n'est pas nécessaire d'exposer toutes vos statistiques descriptives et toutes les variantes des modèles que vous allez imaginer puis tester au cours de vos recherches. Mais vous devrez toutefois expliquer vos différentes tentatives et donc les raisons du choix de la spécification finale du modèle donnant la meilleure prédiction. Vous présenterez évidemment votre modèle final estimé et la qualité de cette estimation.

Enfin vous testerez à l'aveugle les qualités prédictives de votre modélisation sur des données plus récentes. Pour cela vous disposez de deux mois supplémentaires (juin et juillet 2018) de données de l'affluence observée. Ces nouvelles données sont dans le fichier CrowdCounderNewdata.

4 Devoir

Par groupe de 2 étudiants (les binômes devront être composés différemment par rapport au premier devoir portant sur l'hérosécédasticité), **vous restituerez 1) votre script R Markdown, 2) l'output html et 3) une note de synthèse de deux pages en L^AT_EX décrivant votre analyse et vos résultats finaux.**

Le nom des fichiers transmis doit **comporter clairement les noms des étudiants** de chaque groupe.

En cas de difficulté et pour répondre à vos questions, exposez vos questions par email. Au besoin nous échangerons à distance par le moyen le plus adapté.

Ce travail sera évalué. Il est à rendre au cours de la journée du **vendredi 24 avril et avant minuit dernier délai** par simple envoi d'un email à, conjointement, alain.bousquet@univ-tours.fr et yann.kossi@univ-tours.fr